aurigin

Document Summary

New Search Help

Preview Claims
Preview Full Text
Preview Full Image

Email Link:

Document ID: J P 04-018673 A2
Title: METHOD AND DEVICE FOR EXTRACTING TEXT INFORMATION
Assignee: HITACHI LTD
Inventor: MORIMOTO YASUTSUGU
YAMANO FUMIYUKI
US Class:
Int'l Class: G06F 15/401 A
Issue Date: 01/22/1992
Filing Date: 05/11/1990

Abstract:

PURPOSE: To easily detect a part corresponding to an abstract by storing information of relations between sentences in a document and abstracted sentences and obtaining sentences of the original document corresponding to abstracted sentences in accordance with this information and outputting text information such as a list of words appearing in these sentences.

CONSTITUTION: A document is retrieved from an English text file and is subjected to morpheme analysis, and the results are stored in an appearing word table. The range of sentences in the original document corresponding to each abstracted sentence is determined and the abstract is displayed, and text information is extracted and displayed. Thus, a part of original sentences is accurately displayed to efficiently acquire text information.

⑲日本国特許庁（JP）　⑪特許出願公開

⑫公開特許公報（A）　平4－18673

⑤1Int. Cl.[5]　識別記号　庁内整理番号　⑬公開　平成4年(1992)1月22日
G 06 F 15/401　7056－5L

審査請求　未請求　請求項の数 17（全28頁）

⑤4発明の名称　テキスト情報抽出方法および装置

⑵1特　願　平2－122411
⑵2出　願　平2(1990)5月11日

⑦2発明者　森本　康嗣　神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内
⑦2発明者　山野　文行　神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内
⑦1出願人　株式会社日立製作所　東京都千代田区神田駿河台4丁目6番地
⑦4代理人　弁理士　有近　紳志郎

明細書

1．発明の名称

テキスト情報抽出方法および装置

2．特許請求の範囲

1．文書と，その文書の抄録と，前記文書中の文章と前記抄録中の文の関係情報とを記憶する記憶ステップと、

前記抄録中の文のうちの少なくとも1つを選択する抄録文選択ステップと、

その選択された抄録文に対応する前記文書中の文章を前記関係情報を用いて前記文書から抽出する対応文章抽出ステップと、

その抽出された対応文章についてのテキスト情報を取り出して出力するテキスト情報出力ステップと

を有することを特徴とするテキスト情報抽出方法。

2．記憶ステップにおいて、文書中の文と抄録文の1対1の対応情報を記憶し、その記憶した1対1の対応情報から抄録文に対応する文書中の文を調べてその文を含む複数の文を対応文章の範囲として決定し、その対応文章の範囲を関係情報として記憶する請求項1のテキスト情報抽出方法。

3．文書の抄録を自動作成し、その抄録の自動作成の過程で得られた情報から前記文書中の文と抄録文の関係情報を自動作成する請求項1または請求項2のテキスト情報抽出方法。

4．複数の文書の中から1つの文書を選択するための文書選択情報を入力して文書を選択する文書選択ステップをさらに有し、

前記選択した文書の抄録を自動作成し、その抄録の自動作成の過程で得られた情報から前記文書中の文と抄録文の関係情報を自動作成する請求項1または請求項2のテキスト情報抽出方法。

5．ある文書についてのテキスト情報出力ステップ中に、他の文書に関しての抄録文選択ステップと，対応文章抽出ステップと，テキスト情報出力ステップとを再帰的に実行するべ

く、複数の文書の中から前記他の文書を選択する他文書選択ステップをさらに有する請求項４のテキスト情報抽出方法。

６．文書選択情報により選択した文書の抄録を自動作成する際に、前記文書選択情報を利用する請求項４または請求項５のテキスト情報抽出方法。

７．テキスト情報の種類をユーザが選択するテキスト情報選択ステップをさらに有する請求項１から請求項６のいずれかのテキスト情報抽出方法。

８．抄録文に対応する文書中の文を含む段落の全文を対応文章の範囲として決定する請求項１から請求項７のいずれかのテキスト情報抽出方法。

９．抄録文に対応する文書中の文より所定の前方相対値だけ前の文から所定の後方相対値だけ後の文までの全文を対応文章の範囲として決定する請求項１から請求項７のいずれかのテキスト情報抽出方法。

１０．対応文章の最初の語句または最後の語句が所定の語句であるときは、その所定の語句に応じて対応文章の範囲を前後に拡張し、新たな対応文章とする対応文章拡張ステップをさらに有する請求項１から請求項９のいずれかのテキスト情報抽出方法。

１１．対応文章をユーザに提示し、ユーザからの変更指示入力に応じて対応文章の範囲を前後に変更し、新たな対応文章とする対応文章変更ステップをさらに有する請求項１から請求項１０のいずれかのテキスト情報抽出方法。

１２．テキスト情報が、対応文章の範囲内の各文である請求項１から請求項１１のいずれかのテキスト情報抽出方法。

１３．テキスト情報が、対応文章の範囲内に出現する単語のリストである請求項１から請求項１１のいずれかのテキスト情報抽出方法。

１４．文書中の各段落の先頭の文を抽出して抄録とする請求項１から請求項１３のいずれかのテキスト情報抽出方法。

１５．文書と，その文書の抄録と，前記文書中の文章と前記抄録中の文の関係情報とを記憶する記憶手段と、

その記憶手段に記憶した抄録を出力する抄録出力手段と、

その抄録出力手段により出力された抄録中の文のうちの少なくとも１つをユーザが選択する抄録文選択手段と、

その選択された抄録文に対応する前記文書中の文章を前記関係情報を用いて前記文書から抽出する対応文章抽出手段と、

その抽出された対応文章についてのテキスト情報を取り出して出力するテキスト情報出力手段と

を具備してなることを特徴とするテキスト情報抽出装置。

１６．記憶手段が、文書中の文と抄録文の１対１の対応情報を記憶し、その記憶した１対１の対応情報から抄録文に対応する文書中の文を調べてその文を含む複数の文を対応文章の範囲として決定し、その対応文章の範囲を関係情報として記憶する請求項１５のテキスト情報抽出装置。

１７．テキスト情報出力手段が、選択された抄録文とテキスト情報とを対照可能に同時出力する請求項１５または請求項１６のテキスト情報抽出装置。

３．発明の詳細な説明

［産業上の利用分野］

本発明は、テキスト情報抽出方法および装置に関し、さらに詳しくは、ある文書についてのテキスト情報をその文書の抄録を利用して容易に得られるようにしたテキスト情報抽出方法および装置に関する。

［従来の技術］

従来、ある文書における重要箇所とか，その重要箇所に出現する単語のリストとかのテキスト情報を得たい場合、人が文書を解読して重要箇所を取り出したり，その取り出した重要箇所を１つの文書として出現単語リスト自動抽出装置に入力し

て単語リストを得たりしている。

また、従来、抄録を読んでいて元の文書の文章を参照したい場合、人が文書を解読して対応部分を見つけ出している。

［発明が解決しようとする課題］

しかし、人が文書を解読して重要箇所を取り出したり、抄録に対応する部分を見つけ出すのは非常に手間がかかる問題点がある。

そこで、本発明の目的は、ある文書における重要箇所とか，その重要箇所に出現する単語のリストとかのテキスト情報を容易に得られるようにすると共に、抄録から元の文書の文章を容易に参照できるようにするテキスト情報抽出方法を提供することにある。また、そのテキスト情報抽出方法を好適に実施するテキスト情報抽出装置を提供することにある。

［課題を解決するための手段］

本発明は、文書と，その文書の抄録と，前記文書中の文章と前記抄録中の文の関係情報とを記憶する記憶ステップと、前記抄録中の文のうちの少

なくとも１つを選択する抄録文選択ステップと、その選択された抄録文に対応する前記文書中の文章を前記関係情報を用いて前記文書から抽出する対応文章抽出ステップと、その抽出された対応文章についてのテキスト情報を取り出して出力するテキスト情報出力ステップとを有するテキスト情報抽出方法を提供する。

また、本発明は、文書と，その文書の抄録と，前記文書中の文章と前記抄録中の文の関係情報とを記憶する記憶手段と、その記憶手段に記憶した抄録を出力する抄録出力手段と、その抄録出力手段により出力された抄録中の文のうちの少なくとも１つをユーザが選択する抄録文選択手段と、その選択された抄録文に対応する前記文書中の文章を前記関係情報を用いて前記文書から抽出する対応文章抽出手段と、その抽出された対応文章についてのテキスト情報を取り出して出力するテキスト情報出力手段とを具備してなるテキスト情報抽出装置を提供する。

［作用］

本発明のテキスト情報抽出方法および装置では、ある文書中の文とその抄録中の文（抄録文）との関係情報を記憶しておき、抄録文をユーザが選択すると、その抄録文に対応する元の文書の文章を前記関係情報より求め、その元の文書の文章や，その文章に出現する単語のリストなどのテキスト情報を出力する。

そこで、ある文書の重要箇所や，その重要箇所に出現する単語のリストなどを抄録を利用して容易に得られるようになる。

また、抄録文に対応する元の文書の文章を容易に参照できるようになる。

［実施例］

以下、図に示す実施例により本発明を詳細に説明する。なお、これにより本発明が限定されるものではない。

第１１図は、本発明の第１実施例のテキスト検索・表示システム１０００のブロック図である。

図中、１は処理装置、２はディスプレイ装置、

３はキーボード装置、４はメモリ、６は英文テキストファイル、７はストップワードファイル、８は接続語句ファイル、９は辞書ファイルである。

メモリ４は、文書検索プログラム４０１と，抄録作成プログラム４０２と，対応範囲決定プログラム４０３と，テキスト情報抽出プログラム４０５と，出現単語テーブル４０６と，単語頻度テーブル４０７と，文得点テーブル４０８と，抄録文テーブル４０９と，対応範囲テーブル４１０と，表示単語テーブル４１１と，キーワードテーブル４１２とを含んでいる。

英文テキストファイル６は、複数の英文の文書６０と，各文書毎の管理情報テーブル６１とを含んでいる。この管理情報テーブル６１は、第１２図に示すように、文書の各文についての文番号とその文が属する段落の段落番号とを対応付けたものである。

第１図は、上記テキスト検索・表示システム１０００により英文テキストファイル６から文書を検索し、抄録を作成し、その抄録を表示すると共

に、その抄録に対応する元の文章を英文テキストファイル６から抽出して表示する処理のフロー図である。

ステップ１１では、英文テキストファイル６から文書を検索する。検索方法は、テキストデータベース検索システム等で公知の検索方法（例えば、検索キーを入力し、その検索キーと予め文書に付与しておいた文書キーとを比較して、一致した文書を取り出す方法等）を利用できる。

ステップ１２では、前記取り出した文書の各文を形態素解析し、各文の構成単語を出現単語テーブル４０６に出現順に格納する。すなわち、第１３図に示すように、文書６０の各文の単語を切り出して順に出現単語テーブル４０６に格納する。形態素解析の方法は、例えば特開昭５８－４０６８４号公報等に開示の方法を利用できる。

ステップ１３では、前記文書の抄録を作成する。第２図は、抄録作成処理の一例である。

まず、ステップ４１では、文書の文中に出現する単語とその出現頻度を求め、第１４図に示す如

き単語頻度テーブル４０７に格納する。ただし、第１５図に示す如きストップワードファイル７で指定されている単語（これをストップワードという）の出現頻度は求めない。なお、ストップワードファイル７におけるストップワードの指定は、前置詞，冠詞といった品詞での指定と，be，have といった単語での指定の２つの指定方法で行なわれ、前者で指定したときは指定方法ラベルに“０”を設定し、後者で指定したときは指定方法ラベルに“１”を設定し、両者を区別する。

ステップ４２では、予め設定された値以上の出現頻度を持つ単語を重要語として選ぶ。また、前記第１図のステップ１１で文書の検索に検索キーが用いられたときは、その検索キーも重要語に加える。

ステップ４３では、各文における重要語数を各文における単語数で除算して各文の得点を求める。求めた得点は、第１６図に示す如き文得点テーブル４０８に格納する。

ステップ４４では、得点の高い順に文を予め設

定した数だけ取り出し、それらの文（これらを抄録文という）を出現順に並べて抄録とする。実際には、抄録テーブルを作成する。

抄録の作成方法の他の例としては、例えば文書中の各段落の先頭文を取り出し順に並べて抄録とする方法を用いることが出来る。この方法は最も簡単であるため、簡便な抄録で足る場合に有用である。また、例えば「運用Ⅱ・人文研究のための言語データ処理入門／朝倉日本語新講座６ pp. 2-4（朝倉書店刊，1983）」に開示の方法を利用できる。さらに、特開昭６１－１１７６５８公報に開示の方法を利用できる。

第１図に戻り、ステップ４５では、各抄録文に順に抄録文番号を付し、その抄録文番号と元の文番号とを対応づけて、第１７図に示す如き抄録文テーブル４０９を作成する。

ステップ１４では、各抄録文に対応する元の文書中の文章の範囲を決定する。

この対応範囲決定処理を第３図を参照して説明する。

ステップ９１では、抄録文番号ｉ（A_i）を“１”（Ａ１）に初期設定する。

ステップ９２では、抄録文番号ｉ（A_i）に対応する元の文番号ｊ（B_j）を第１７図の抄録文テーブル４０９から求める。ｉ＝１（A_1）なら、ｊ＝２（B_2）が求まる。

ステップ９３では、文番号ｊ（B_j）の文が属する段落の段落番号ｋ（D_k）を第１２図の管理情報テーブル６１から求める。ｊ＝２（B_2）なら、ｋ＝１（D_k）が求まる。

ステップ９４では、段落番号ｋ（D_k）の段落の最初の文の文番号を前記管理情報テーブル６１から求めて、抄録文番号ｉ（A_i）の抄録文に対応する元の文書中の文章の範囲の開始位置とする。

ステップ９５では、段落番号ｋ（D_k）の段落の最後の文の文番号を前記管理情報テーブル６１から求めて、抄録文番号ｉ（A_i）の抄録文に対応する元の文書中の文章の範囲の終了位置とする。

ステップ９６では、前記抄録文番号ｉ（A_i）の抄録文に対応する文章の範囲を拡張する。

この対応範囲拡張処理を第４図を参照して説明する。

ステップ１１１では、前記開始位置の文が、接続語句ファイル８に登録されている接続語句を含んでいるかどうか調べる。接続語句を含んでおればステップ１１２に進み、含んでいなければステップ１１４に進む。

接続語句ファイル８には、第１８図に示すように、接続語句（接続詞やそれに準ずる語句）とその接続語句の性質に応じた対応文章範囲の拡張幅とが格納されている。例えば接続語句“but”とそれに応じた拡張幅“－１”とが格納されている。これは接続語句“but”を含む文が開始位置の文なら、その前の文から読む必要があることを意味している。

ステップ１１２では、拡張幅が負か調べる。負ならステップ１１３に進み、負でないならステップ１１４に進む。

ステップ１１３では、（開始位置＋拡張幅）を新たな開始位置とする。例えば開始位置の文が接

続語句“but”を含んでいたなら、開始位置が１文前に拡張される。

ステップ１１４では、前記終了位置の文が、接続語句ファイル８に登録されている接続語句を含んでいるかどうか調べる。接続語句を含んでおればステップ１１５に進み、含んでいなければ対応範囲拡張処理を終了する。

ステップ１１５では、拡張幅が正か調べる。正ならステップ１１６に進み、正でないなら対応範囲拡張処理を終了する。

ステップ１１６では、（終了位置＋拡張幅）を新たな終了位置とする。例えば終了位置の文が接続語句“as follows”を含んでいたなら、終了位置が１文後に拡張される。そして、対応範囲拡張処理を終了する。

第３図に戻り、ステップ９７では、抄録文とその対応範囲を第１９図に示す如き対応範囲テーブル４１０に登録する。

ステップ９８では、抄録文番号ｉをインクリメントする。

ステップ９９では、抄録文番号ｉが抄録文テーブル４０９の最後を過ぎたか調べる。過ぎていなければ前記ステップ９２に戻り、過ぎていれば対応範囲決定処理を終了する。

第１図に戻り、ステップ１５では、抄録を表示する。表示例を第２２図に示す。

ステップ１６では、テキスト情報を表示することをユーザが選択したかチェックする。選択したならステップ１７に進み、選択しなければ処理を終了する。

ステップ１７では、テキスト情報を抽出し、表示する。

このテキスト情報抽出・表示処理を第５図を参照して説明する。

ステップ１５１では、表示された抄録の中からテキスト情報を得たい抄録文をユーザが選択する。

ステップ１５２では、選択された抄録文とその前後のいくつかの抄録文に対応する文章の範囲を第１９図の対応範囲テーブル４１０から求めて、それぞれ処理範囲とする。

ステップ１５３では、処理モードをユーザが選択する。処理モードには、原文表示処理と，キーワード検索処理と，他文書検索処理と，抄録文選択処理と，終了とがあり、これらのいずれかを選択すると、ステップ１５４～ステップ１５７のいずれかに進むか又はテキスト情報抽出・表示処理を終了する。

ステップ１５４では、原文表示処理を行なう。この原文表示処理を第６図を参照して説明する。

ステップ１６１では、抄録文と前記処理範囲の文章を対照表示するか否かをユーザに選択させる。対照表示を選択するとステップ１６２に進み、選択しなければステップ１６５に進む。

ステップ１６２では、画面を左右に分割する。

ステップ１６３では、前記第５図のステップ１５１でユーザが選択した抄録文を含む抄録の部分を画面の左側エリアに表示する。この際、ユーザが選択した抄録文が左側エリアの中心の行に位置するように、また、輝度を強くするなどにより他と識別可能に表示する。さらに、前記左側エリア

に表示した抄録文に対応する処理範囲の文章を英文テキストファイル６の文書６０から読み込んで、画面の右側エリアに表示する。前記ユーザが選択した抄録文に対応する文章中のユーザが選択した文は、表示輝度を強くするなどにより他と識別可能に、また、ユーザが選択した文が右側のエリアの中心の行に位置するように表示する。このときの画面の一例を第２３図に示す。

一方、ステップ１６５では、前記処理範囲の文章を英文テキストファイル６の文書６０から読み込んで、画面に表示する。この際、ユーザが選択した抄録文に対応する文章中の文が画面の中心の行に位置するように、また、輝度を強くするなどにより他と識別可能に表示する。このときの画面の一例を第２４図に示す。

ステップ１６７では、ユーザが選択した抄録文に対応する文章が適正か否かをユーザが判断し、適正なら処理範囲を変更しないと入力し、適正でないなら処理範囲を変更すると入力する。処理範囲を変更しないと入力されたら原文表示処理を終

了する。処理範囲を変更すると入力されたらステップ１６８に進む。

ステップ１６８では、ユーザの指示に基づいて処理範囲の開始位置と終了位置を変更する。処理範囲の変更は、第１９図の対応範囲テーブル４１０の対応範囲の開始位置と終了位置をユーザの指示に従って書き換えることによって行う。そして、前記ステップ１６１に戻る。

第５図に戻り、ステップ１５５では、キーワード検索処理を行なう。このキーワード検索処理を第７図を参照して説明する。

ステップ１９１では、キーワードを入力するための参考とする単語リストを表示するか否かをユーザに選択させる。表示するならばステップ１９２に進み、表示しないならばステップ１９３に進む。

ステップ１９２では、単語リストを表示する。この単語リスト表示処理を第８図を参照して説明する。

ステップ２０１では、表示する単語リストの種

類をユーザに選択させる。表示する単語リストの種類には、出現頻度順単語リストおよびアルファベット順単語リストの２種類がある。

ステップ２０２では、前記処理範囲の文章に出現する単語を第１３図の出現単語テーブル４０６から得て、第２０図の表示単語テーブル４１１に格納する。但し、ストップワードファイル７で指定されている単語は除く。また、この時点では、表示単語テーブル４１１の各単語の出現頻度の欄は、初期値として“１”を設定しておく。

ステップ２０３では、表示単語テーブル４１１中に重複して出現する単語を探し、その数をカウントし、その単語を含むレコードを一つだけ残して他を削除し、残したレコードの出現頻度を前記カウントした数で置き換える。

ステップ２０４では、前記ステップ２０１で出現頻度順単語リストが指定されたか否かを調べる。指定されている場合にはステップ２０５に進み、指定されていない場合にはステップ２０６に進む。

ステップ２０５では、表示単語テーブル４１１を出現頻度順にソートする。

一方、ステップ２０６では、表示単語テーブル４１１をアルファベット順にソートする。

ステップ２０７では、表示単語テーブル４１１をリスト化して表示する。第２５（ａ）図に出現頻度順単語リストを例示する。また、第２５（ｂ）図にアルファベット順単語リストを例示する。

第７図に戻り、ステップ１９３では、キーワードをユーザが入力する。この入力は、前記ステップ１９２で表示された単語リスト中の単語を選択することにより入力してもよいし、キーボード装置３から入力してもよい。また、複数のキーワードを入力してもよい。キーワードが複数ある場合には、処理が繰返しになるだけである。入力されたキーワードは、第２１図に示す如きキーワードテーブル４１２に格納する。

ステップ１９４では、前記キーワードテーブル４１２に格納したキーワードを含む文を処理範囲内の文章から抽出する。キーワードを含む文の文番号は、第１３図の出現単語テーブル４０６を調

べることによって得られる。

ステップ１９５では、前記抽出したキーワードを含む文を英文テキストファイル６の文書６０から取り出して、表示する。表示方法としては、第２６（ａ）図に示すように、キーワードを含む文のみを表示してもよいし、第２６（ｂ）図に示すように、処理範囲の文章を表示しておき、キーワードを含む文をブリンク等により強調してもよい。

ステップ１９６では、第５図のステップ１５１でユーザが選択した抄録文に対応する処理範囲が適正か否かをユーザが判断し、適正なら処理範囲を変更しないと入力し、適正でないなら処理範囲を変更すると入力する。処理範囲を変更しないと入力されたらキーワード検索処理を終了する。処理範囲を変更すると入力されたらステップ１９７に進む。

ステップ１９７では、ユーザの指示に基づいて処理範囲の開始位置と終了位置を変更する。処理範囲を変更すると、それにより第１９図の対応範囲テーブル４１０の対応範囲の開始位置と終了位

置が更新される。そして、前記ステップ１９４に戻る。

第５図に戻り、ステップ１５６では、他文書検索処理を行なう。この他文書検索処理を第９図を参照して説明する。

ステップ２５１では、キーワードを入力する参考とするために単語リストを表示するか否かをユーザに選択させる。表示するのならば、ステップ２５２に進む。表示しないのならば、ステップ２５３に進む。このステップ２５１は、前記第７図のステップ１９１と同様である。

ステップ２５２では、単語リストを表示する。このステップ２５２は、前記第７図のステップ１９２と同様であり、具体的には第８図の処理となる。

ステップ２５３では、キーワードを入力する。このステップ２５３は、前記第７図のステップ１９３と同様である。

この後、第５図に戻ってから第１図のステップ１１へ移行する。そして、キーワードテーブル４

１２中の単語を検索キーとして他の文書を検索し、更にステップ１２からステップ１７の一連の処理を再帰的に行う。

この他文書検索処理によって、ある文書のテキスト情報を抽出している際に、関連する他の文書を連想的に検索できる。すなわち、ある文書中の単語や，ある文書を読んでいて思い付いた単語を検索キーとして、他の文書を連想的に検索できる。

さて、第５図に戻り、ステップ１５７の抄録文選択処理では、抄録を表示し、前記ステップ１５１に戻る。

以上の実施例では英文テキストファイル６の管理情報テーブル６１に段落の情報を持っていたが、次に変形実施例として段落の情報を持たない場合を説明する。

この変形実施例の場合、第１図のステップ１４の処理として、前記第３図のステップ９１～ステップ９９を行なう代わりに、第１０図のステップ２６１～ステップ２６９を行なう。

すなわち、ステップ２６１では、ある抄録文に

対応する元の文書６０の文を基準とした相対的な値で、その抄録文に対応する元の文書６０の文章の範囲をユーザが設定する。具体的には、ある抄録文に対応する元の文書６０の文を“０”とし、その文より前の文は負の値．その文より後の文は正の値で文の数をカウントして、対応範囲の相対的開始位置αと相対的終了位置βとを設定する。この設定は、予め行なっていてもよい。

ステップ２６２では、抄録文番号ｉ（A_i）を“１”に初期化する。

ステップ２６３では、抄録文番号ｉ（A_i）の抄録文の元の文番号ｊ（B_i）を第１７図の抄録文テーブル４０９から求める。

ステップ２６４では、文番号（ｊ＋α）の文を、抄録文番号ｉ（A_i）の抄録文の対応範囲の開始位置とする。

ステップ２６５では、文番号（ｊ＋β）の文を、抄録文番号ｉ（A_i）の抄録文の対応範囲の終了位置とする。

ステップ２６６では、第１８図の接続語句ファ

イル８を参照して対応範囲を拡張する。この処理は第３図のステップ９６と同様であり、具体的には第４図に示す処理となる。

ステップ２６７では、抄録文とその対応範囲を第１９図に示す如き対応範囲テーブル４１０に登録する。

ステップ２６８では、抄録文番号ｉをインクリメントする。

ステップ２６９では、抄録文番号ｉが抄録文テーブル４０９の最後を過ぎたか調べる。過ぎていなければ前記ステップ２６３に戻り、過ぎていれば対応範囲決定処理を終了する。

他の変形実施例としては、第１図のステップ１１～ステップ４５を、ユーザが抄録を作成して入力するステップと，ユーザが第１７図の抄録文テーブル４０９を作成して入力するステップとに置換したものが挙げられる。

上記実施例のテキスト情報抽出方法および装置によれば、抄録中で分かりにくかったり，特に関心が持たれた抄録文に関係する元の文章の一部を

的確に表示したり、出現単語リストを表示することにより、必要十分なテキスト情報を効率的に獲得することが出来る。

以上の第１実施例は抄録文に対応する元の文章を検索して表示するシステムであったが、次に第２実施例として抄録文とその翻訳文とを組み合わせて表示するシステムについて説明する。

第３３図は、本発明の第２実施例のテキスト検索・表示システム２０００のブロック図である。

図中、１は処理装置、２はディスプレイ装置、３はキーボード装置、４はメモリ、５は日本文テキストファイル、６は英文テキストファイル、７はストップワードファイル、８は接続語句ファイル、９は辞書ファイルである。

メモリ４は、文書検索プログラム４０１と，抄録作成プログラム４０２と，対応範囲決定プログラム４０３と，自動翻訳プログラム４０４と，テキスト情報抽出プログラム４０５と，出現単語テーブル４０６と，単語頻度テーブル４０７と，文得点テーブル４０８と，抄録文テーブル４０９と，

対応範囲テーブル４１０と，表示単語テーブル２４１１と，キーワードテーブル４１２とを含んでいる。

英文テキストファイル６は、複数の英文の文書６０と，各文書毎の管理情報テーブル６１とを含んでいる。この管理情報テーブル６１は、第１２図に示すように、文書の各文についての文番号とその文が属する段落の段落番号とを対応付けたものである。

第２７図は、上記テキスト検索・表示システム２０００において英文テキストファイル６から文書を検索し、抄録を作成し、その抄録を表示すると共に、その抄録を翻訳し、抄録と対応付けて表示する処理のフロー図である。

ステップ２７１では、英文テキストファイル６から文書を検索する。この処理は第１図のステップ１１と同様である。

ステップ２７２では、前記取り出した文書の文を形態素解析し、各文の構成単語を第１３図の出現単語テーブル４０６に出現順に格納する。この

処理は第１図のステップ１２と同様である。

ステップ２７３では、前記文書の抄録を作成する。この処理は第１図のステップ１３と同様である。

ステップ２７４５では、各抄録文に順に抄録文番号を付し、その抄録文番号と元の文番号とを対応づけて、第１７図に示す如き抄録文テーブル４０９を作成する。この処理は第１図のステップ４５と同様である。

ステップ２７４では、各抄録文に対応する元の文書中の文章の範囲を決定する。この処理は第１図のステップ１４と同様である。

ステップ２７５では、抄録文に対し翻訳処理を行う。翻訳処理は、例えば特開昭５８－４０６８４号公報に開示の翻訳方法を利用することが出来る。翻訳文は、抄録文の元の文番号と対応付けて日本文テキストファイル５に格納する。英文テキストファイル６の文書６０のデータ構造を第３４(ａ)図に示し、日本文テキストファイル５のデータ構造を第３４(ｂ)図に示す。

ステップ２７６では、抄録を表示する。その際、ユーザは、表示言語選択キーにより抄録文を表示する言語として英語または日本語を選択する。英語が選択されたら、英文テキストファイル６の文書６０から抄録文になっている文を抽出して表示する。この英語での表示例を第３６(ａ)図に示す。日本語が選択されたら、日本文テキストファイル５から翻訳文を抽出して表示する。この日本語での表示例を第３６(ｂ)図に示す。

ステップ２７７では、抄録を読んだユーザが、テキスト情報を表示するか否か入力する。この処理は、第１図のステップ１６と同様である。表示するならステップ２７８に進み、表示しないなら処理を終了する。

ステップ２７８では、テキスト情報を抽出し、表示する。このテキスト情報抽出・表示処理を第２８図を参照して説明する。

ステップ３０１では、表示された抄録の中からテキスト情報を得たい抄録文をユーザが選択する。

ステップ３０２では、選択された抄録文とその

前後のいくつかの抄録文に対応する文章の範囲を第１９図の対応範囲テーブル４１０から求めて、それぞれ処理範囲とする。

ステップ３０３では、処理範囲の文章に翻訳処理を行なう。翻訳文は、日本文テキストファイル５に格納する。また、翻訳処理で得られた英語出現単語と訳語を第３５図に示す如き表示単語テーブル２４１１に格納する。この時点では、出現頻度の欄は初期値として“１”を設定しておく。

ステップ３０４では、処理モードをユーザが選択する。処理モードには、原文表示処理と，キーワード検索処理と，他文書検索処理と，抄録文選択処理と，終了とがあり、これらのいずれかを選択すると、ステップ３０５～ステップ３０８のいずれかに進むか又はテキスト情報抽出・表示処理を終了する。

ステップ３０５では、原文表示処理を行なう。この原文表示処理は、第６図を参照して先述した処理と同様である。但し、ユーザは、表示言語選択キーにより原文を表示する言語として英語また

は日本語を選択することが出来る。英語が選択されたら、英文テキストファイル６の文書６０から処理範囲の文を抽出して表示する。この場合の処理範囲の文のみの表示例は第２４図のようになる。また、抄録文と処理範囲の文の対照表示の例は第２３図のようになる。一方、日本語が選択されたら、日本文テキストファイル５から前記処理範囲に対応する翻訳文を抽出して表示する。この場合の処理範囲に対応する翻訳文のみの表示例を第３７(ａ)図に示す。また、抄録文の翻訳文と処理範囲の翻訳文の対照表示の例を第３７(ｂ)図に示す。

ステップ３０６では、キーワード検索処理を行なう。このキーワード検索処理を第２９図を参照して説明する。

ステップ３４１では、テキスト情報を検索するためのキーワードを設定するための参考とする単語リストを表示するか否かをユーザに選択させる。表示するならばステップ３４２に進み、表示しないならばステップ３４３に進む。

ステップ３４２では、単語リストを表示する。

この単語リスト表示処理を第３０図を参照して説明する。

ステップ３５０１では、表示する単語リストの種類をユーザに選択させる。表示する単語リストの種類は、出現頻度順単語リストおよびアルファベット／５０音順単語リストの２種類がある。

ステップ３５０２では、前記第２８図のステップ３０３で作成した第３５図の表示単語テーブル２４１１中で英語出現単語と訳語の両方が一致する単語を探し、複数個ある場合にはその数をカウントし、そのレコードを１つ残して他を削除し、残したレコードの出現頻度の欄にカウントした数を設定する。

ステップ３５０３では、単語を表示する言語として英語または日本語を表示言語選択キーによりユーザに選択させる。英語が選択されたら、ステップ３５０４に進む。日本語が選択されたら、ステップ３５０８に進む。

ステップ３５０４では、前記ステップ３５０１で出現頻度順単語リストが選択されたか否かを調

べ、選択されている場合はステップ３５０５に進み、選択されていない場合はステップ３５０６に進む。

ステップ３５０５では、第３５図の表示単語テーブル２４１１の英語出現単語を出現頻度順にソートする。

ステップ３５０６では、第３５図の表示単語テーブル４１１の英語出現単語をアルファベット順にソートする。

ステップ３５０７では、ソートした表示単語テーブル２４１１の英語出現単語を表示する。第３８(ａ)図に英語出現単語の出現頻度順単語リストを例示する。また、第３８(ｂ)図に英語出現単語のアルファベット順単語リストを例示する。

一方、ステップ３５０８では、前記ステップ３５０１で出現頻度順単語リストが選択されたか否かを調べ、選択されている場合はステップ３５０９に進み、選択されていない場合はステップ３５１０に進む。

ステップ３５０９では、第３５図の表示単語テーブル２４１１の訳語を出現頻度順にソートする。

ステップ３５１０では、第３５図の表示単語テーブル４１１の訳語を５０音順にソートする。

ステップ３５１１では、ソートした表示単語テーブル２４１１の訳語を表示する。第３９(ａ)図に訳語の出現頻度順単語リストを例示する。また、第３９(ｂ)図に訳語の５０音順単語リストを例示する。

第２９図に戻り、ステップ３４３では、キーワードを設定する。このキーワード設定処理を第３１図を参照して説明する。

ステップ３８１では、ユーザがキーワードを入力する。キーワードは、英語，日本語どちらでもよく、複数でもよい。

ステップ３８２では、ユーザが入力したキーワードが英語か否かを識別する。具体的には、キーワードがキーボード装置３から入力された場合にはキーボード装置３の入力モードから英語か否かを識別する。また、表示している単語リストから選択された場合には、その単語リストの種類から

英語か否かを識別する。キーワードが英語の場合にはステップ３８３に進み、日本語の場合にはステップ３８４に進む。

ステップ３８３では、英語のキーワードをキーワードテーブル４１２に格納する。

一方、ステップ３８４では、日本語のキーワードと同じ訳語を持つレコードを第３５図の表示単語テーブル２４１１で探す。

ステップ３８５では、日本語のキーワードと同じ訳語を持つレコードが表示単語テーブル２４１１にあったかチェックする。あったならばステップ３８６に進む。なかったならば、前記ステップ３８１に戻り、キーワードを再入力させる。

ステップ３８６では、日本語のキーワードと同じ訳語を持つレコードの英語出現単語をキーワードテーブル４１２にキーワードとして格納する。もし、日本語のキーワードと同じ訳語を持つレコードが複数ある場合には、その全ての英語出現単語をキーワードテーブル４１２にキーワードとして格納する。

第２９図に戻り、ステップ３４４では、前記キーワードテーブル４１２に格納したキーワードを含む文の文番号を処理範囲内の文章から抽出する。この処理は前記第７図のステップ１９４と同様である。

ステップ３４５では、前記抽出したキーワードを含む文を英文テキストファイル６の文書６０から取り出して、表示する。この処理は前記第７図のステップ１９５と同様である。

ステップ３４６では、処理範囲を変更するか否かをユーザに入力させ、処理範囲を変更しないと入力されたらキーワード検索処理を終了する。処理範囲を変更すると入力されたらステップ３４７に進む。この処理は前記第７図のステップ１９６と同様である。

ステップ３４７では、ユーザの指示に基づいて処理範囲の開始位置と終了位置を変更する。処理範囲を変更すると、それにより第１９図の対応範囲テーブル４１０の対応範囲の開始位置と終了位置が更新される。そして、前記ステップ３４４に

戻る。

さて、第２８図に戻り、ステップ３０７では、他文書検索処理を行なう。この他文書検索処理を第３２図を参照して説明する。

ステップ４２５１では、キーワードを設定する参考とするために単語リストを表示するか否かをユーザに選択させる。表示するのならば、ステップ４２５２に進む。表示しないのならば、ステップ４２５３に進む。このステップ４２５１は、前記第２９図のステップ３４１と同様である。

ステップ４２５２では、単語リストを表示する。このステップ４２５２は、前記第２９図のステップ３４２と同様であり、具体的には第３０図の処理となる。

ステップ４２５３では、キーワードを設定する。このステップ４２５３は、前記第２９図のステップ３４３と同様であり、具体的には第３１図の処理となる。

この後、第２８図に戻ってから第２７図のステップ２７１へ移行する。そして、キーワードテーブル４１２中の単語を検索キーとして他の文書を検索し、更にステップ２７２からステップ２７８の一連の処理を再帰的に行う。

第２８図に戻り、ステップ３０８では、抄録文選択処理を行なう。すなわち、抄録を表示し、前記ステップ３０１に戻る。

変形実施例としては、第２７図のステップ２７１～ステップ２７４５を、ユーザが抄録を作成して入力するステップと，ユーザが第１７図の抄録文テーブル４０９を作成して入力するステップとに置換したものが挙げられる。

他の実施例としては、英語や日本語以外の他の言語について本発明を適用したものが挙げられる。

［発明の効果］

本発明のテキスト情報抽出方法および装置によれば、ある文書における重要箇所とか，その重要箇所に出現する単語のリストとかのテキスト情報を、抄録を利用して容易に得られるようになる。

４．図面の簡単な説明

第１図は本発明の第１実施例の処理の基本フロー図、第２図は抄録作成処理のフロー図、第３図は対応範囲決定処理のフロー図、第４図は対応範囲拡張処理のフロー図、第５図はテキスト情報抽出・表示処理のフロー図、第６図は原文表示処理のフロー図、第７図はキーワード検索処理のフロー図、第８図は単語リスト表示処理のフロー図、第９図は他文書検索処理のフロー図、第１０図は対応範囲決定処理の他の例のフロー図、第１１図は本発明の第１実施例のハードウェア構成図、第１２図は情報管理テーブルの概念図、第１３図は出現単語テーブルの概念図、第１４図は単語頻度テーブルの概念図、第１５図はストップワードファイルの概念図、第１６図は文得点テーブルの概念図、第１７図は抄録文テーブルの概念図、第１８図は接続語句ファイルの概念図、第１９図は対応範囲テーブルの概念図、第２０図は表示単語テーブルの概念図、第２１図はキーワードテーブルの概念図、第２２図，第２３図，第２４図，第２５(ａ)図，第２５(ｂ)図，第２６(ａ)図および第２６(ｂ)図は表示画面の例示図、第２７図は本発明の第２実施例の処理の基本フロー図、第２８図はテキスト情報抽出・表示処理のフロー図、第２９図はキーワード検索処理のフロー図、第３０図は単語リスト表示処理のフロー図、第３１図はキーワード設定処理のフロー図、第３２図は他文書検索処理のフロー図、第３３図は本発明の第２実施例のハードウェア構成図、第３４(ａ)図は英文の文書の概念図、第３４(ｂ)図は日本文テキストファイルの概念図、第３５図は表示単語テーブルの概念図、第３６(ａ)図，第３６(ｂ)図，第３７(ａ)図，第３７(ｂ)図，第３８(ａ)図，第３８(ｂ)図，第３９(ａ)図および第３９(ｂ)図は表示画面の例示図である。

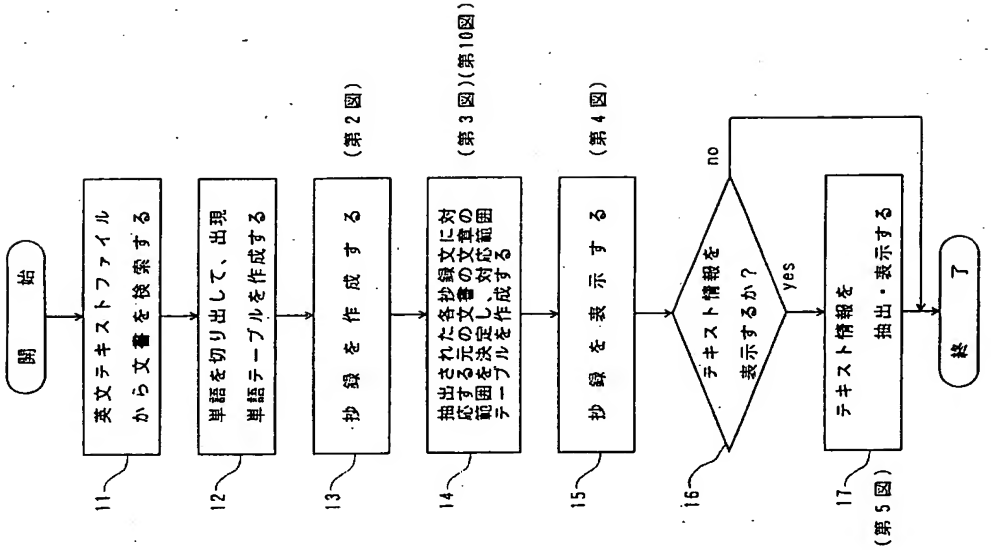
（符号の説明）

１・・・処理装置、２・・・ディスプレイ装置、３・・・キーボード装置、４・・・メモリ、４０１・・・文書検索プログラム、４０２・・・抄録作成プログラム、４０３・・・対応範囲決定プログラム、４０４・・・自動翻訳プログラム、４０５・・・テキスト情報抽出プログラム、４０６・

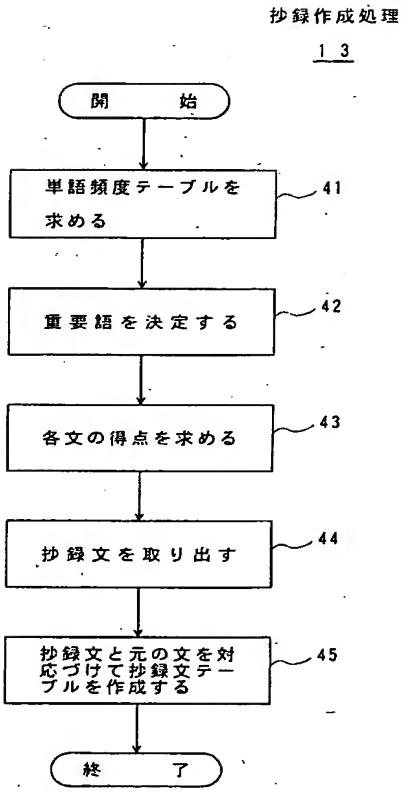
・・出現単語テーブル、４０７・・・単語頻度テーブル、４０８・・・文得点テーブル、４０９・・・抄録文テーブル、４１０・・・対応範囲テーブル、４１１・・・表示単語テーブル、４１２・・・キーワードテーブル、５・・・日本文テキストファイル、６・・・英文テキストファイル、６０・・・文書、６１・・・管理情報テーブル、７・・・ストップワードファイル、８・・・接続語句ファイル、９・・・辞書ファイル。

出願人　株式会社 日立製作所
代理人　弁理士　有近 紳志郎

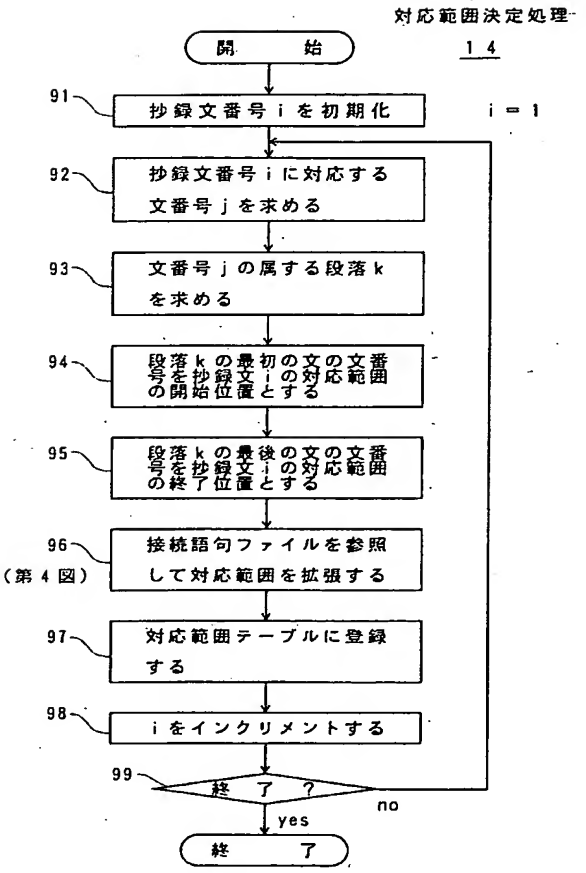
第１図



第 2 図

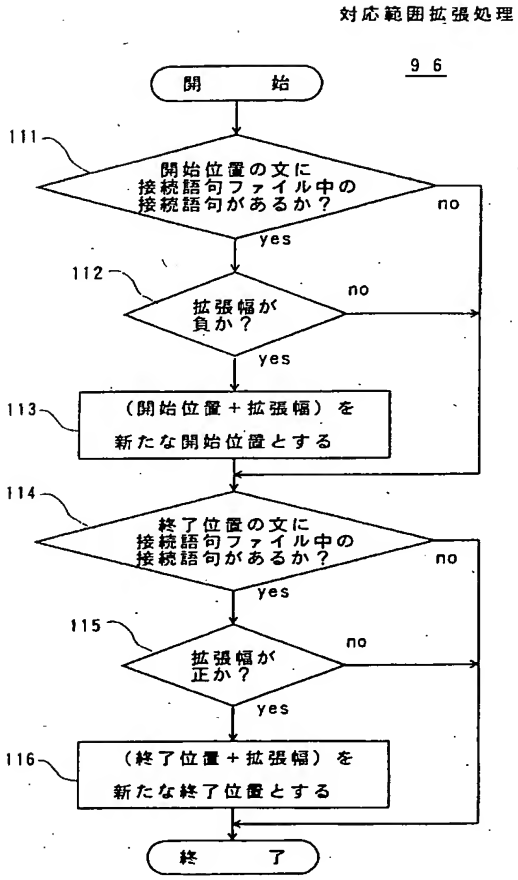

抄録作成処理
13
開　始
単語頻度テーブルを求める
41
重要語を決定する
42
各文の得点を求める
43
抄録文を取り出す
44
抄録文と元の文を対応づけて抄録文テーブルを作成する
45
終　了

第 3 図

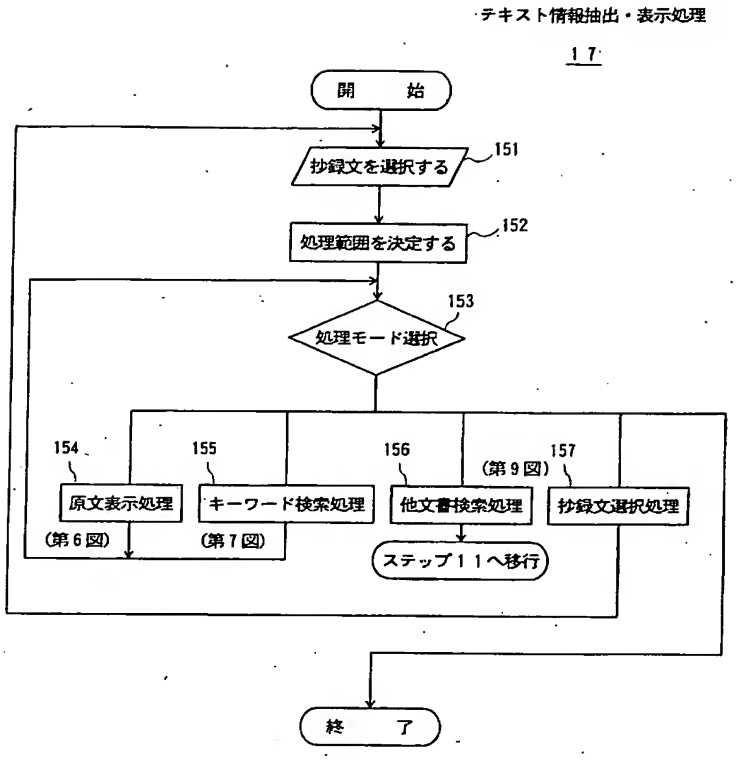

対応範囲決定処理
14
開　始
91
抄録文番号 i を初期化
i = 1
92
抄録文番号 i に対応する文番号 j を求める
93
文番号 j の属する段落 k を求める
94
段落 k の最初の文の文番号を抄録文 i の対応範囲の開始位置とする
95
段落 k の最後の文の文番号を抄録文 i の対応範囲の終了位置とする
96
接続語句ファイルを参照して対応範囲を拡張する
97
対応範囲テーブルに登録する
98
i をインクリメントする
99
終　了　?
no
yes
終　了

(第 4 図)

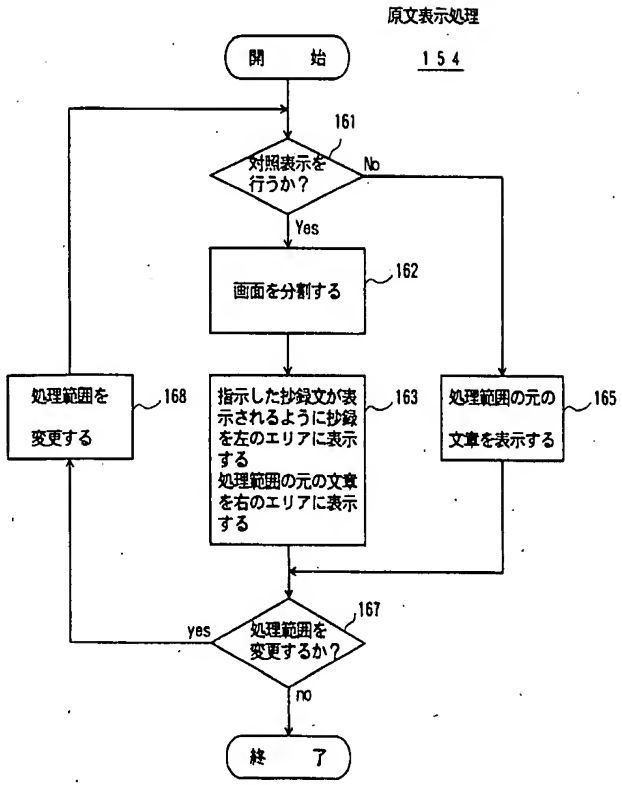
第 4 図


対応範囲拡張処理
96
開　始
111
開始位置の文に
接続語句ファイル中の
接続語句があるか？
no
yes
112
拡張幅が
負か？
no
yes
113
（開始位置＋拡張幅）を
新たな開始位置とする
114
終了位置の文に
接続語句ファイル中の
接続語句があるか？
no
yes
115
拡張幅が
正か？
no
yes
116
（終了位置＋拡張幅）を
新たな終了位置とする
終　了

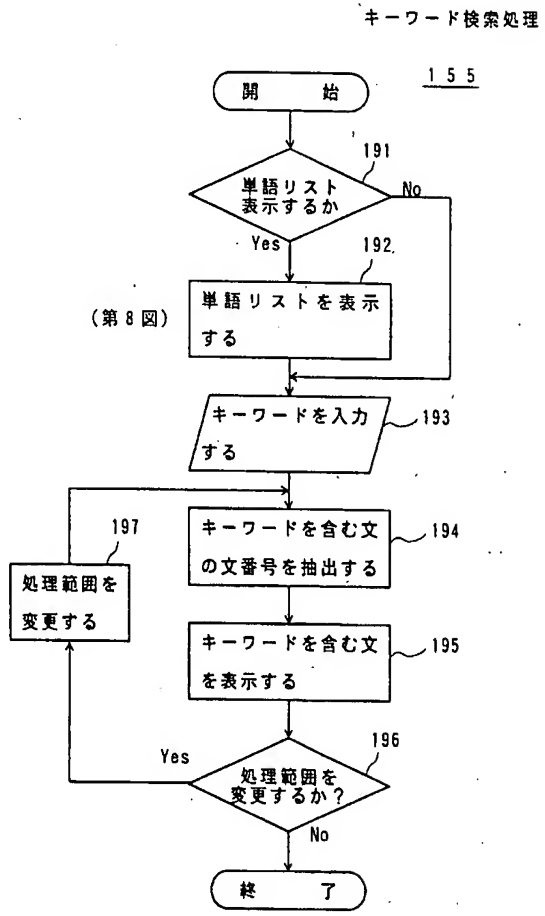
第 5 図


テキスト情報抽出・表示処理
17
開　始
抄録文を選択する
151
処理範囲を決定する
152
153
処理モード選択
154
原文表示処理
(第6図)
155
キーワード検索処理
(第7図)
156
他文書検索処理
(第9図)
157
抄録文選択処理
ステップ11へ移行
終　了

第 6 図

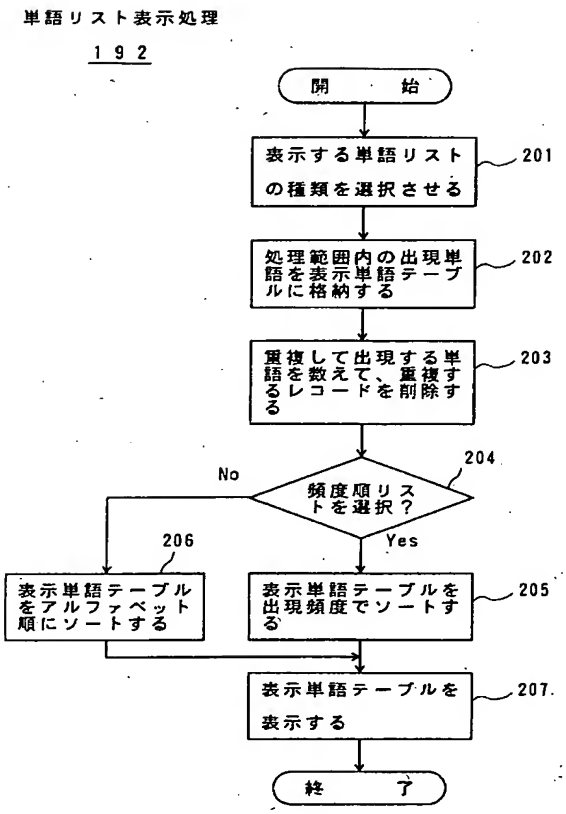

原文表示処理
開　始
154
161
対照表示を行うか？
No
Yes
画面を分割する
162
処理範囲を変更する
168
指示した抄録文が表示されるように抄録を左のエリアに表示する
処理範囲の元の文章を右のエリアに表示する
163
処理範囲の元の文章を表示する
165
yes
処理範囲を変更するか？
167
no
終　了

第 7 図

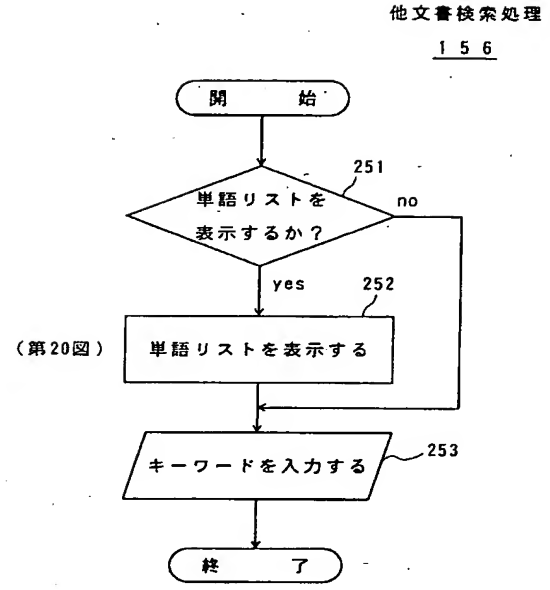

キーワード検索処理
開　始
155
191
単語リスト表示するか
No
Yes
192
単語リストを表示する
キーワードを入力する
193
キーワードを含む文の文番号を抽出する
194
キーワードを含む文を表示する
195
196
処理範囲を変更するか？
Yes
No
197
処理範囲を変更する
終　了

（第 8 図）

第 8 図

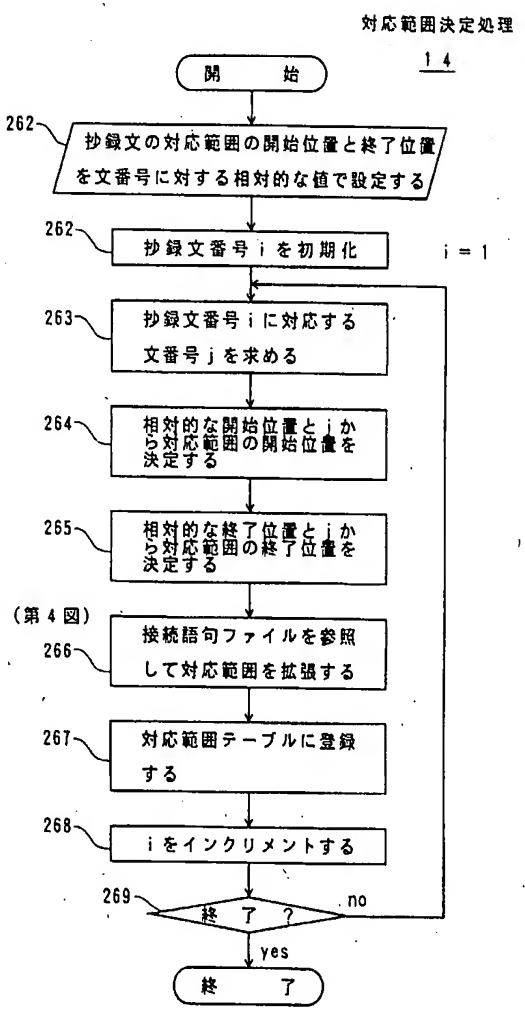

単語リスト表示処理
192
開　始
表示する単語リストの種類を選択させる
201
処理範囲内の出現単語を表示単語テーブルに格納する
202
重複して出現する単語を数えて、重複するレコードを削除する
203
頻度順リストを選択？
204
No
Yes
表示単語テーブルをアルファベット順にソートする
206
表示単語テーブルを出現頻度でソートする
205
表示単語テーブルを表示する
207
終　了

第 9 図

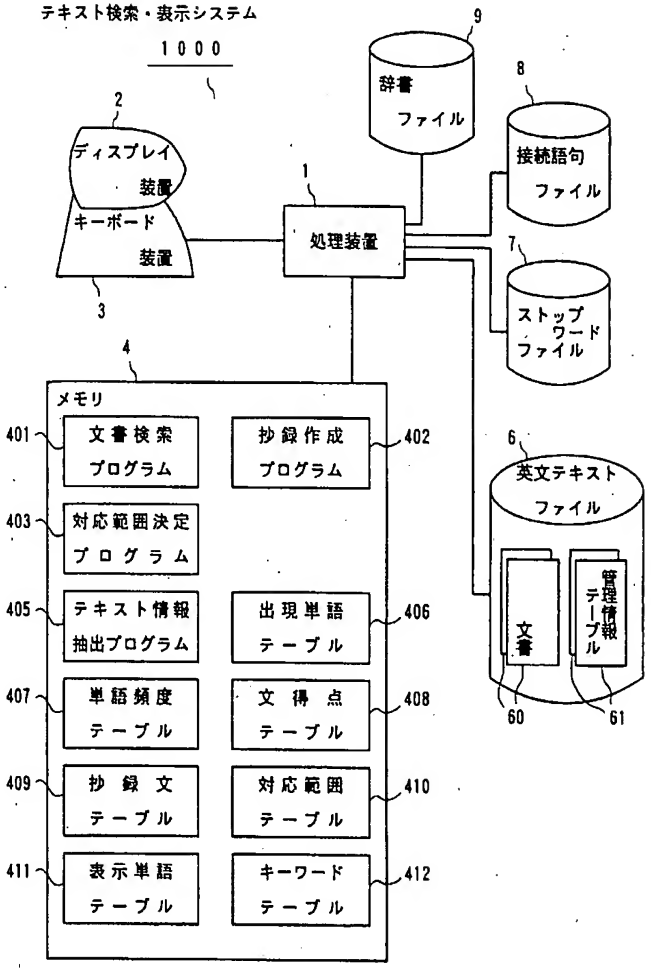

他文書検索処理
156
開　始
単語リストを表示するか？
251
no
yes
252
単語リストを表示する
キーワードを入力する
253
終　了

（第20図）

第 10 図


対応範囲決定処理
14
開　始
262
抄録文の対応範囲の開始位置と終了位置を文番号に対する相対的な値で設定する
262
抄録文番号 i を初期化
i = 1
263
抄録文番号 i に対応する文番号 j を求める
264
相対的な開始位置と j から対応範囲の開始位置を決定する
265
相対的な終了位置と j から対応範囲の終了位置を決定する
(第 4 図)
266
接続語句ファイルを参照して対応範囲を拡張する
267
対応範囲テーブルに登録する
268
i をインクリメントする
269
終　了　?
no
yes
終　了

第 11 図


テキスト検索・表示システム
1000
2
ディスプレイ装置
キーボード装置
3
1
処理装置
9
辞書ファイル
8
接続語句ファイル
7
ストップワードファイル
6
英文テキストファイル
文書
管理情報テーブル
60
61
4
メモリ
401
文書検索プログラム
402
抄録作成プログラム
403
対応範囲決定プログラム
405
テキスト情報抽出プログラム
406
出現単語テーブル
407
単語頻度テーブル
408
文得点テーブル
409
抄録文テーブル
410
対応範囲テーブル
411
表示単語テーブル
412
キーワードテーブル

第 12 図

管理情報テーブル

61

文 番 号	段落番号
B 1	D 1
B 2	D 1
B 3	D 1
B 4	D 2
・	・
・	・
・	・

第 13 図

出現単語テーブル

406

文番号	出 現 単 語			
B 1	Japan	role	in	・・・
B 2	This	movement	have	・・・
B 3	The	number	of	・・・
B 4	it	be	very	・・・
・	・	・	・	
・	・	・	・	
・	・	・	・	

文書

60

文番号	英 文 デ ー タ
B 1	Japan's role in the world has expanded dramatically.
B 2	This movement has necessitated translation of ・・・
B 3	The number of translators in the labor market ・・・
B 4	It is very ・・・

第 15 図

ストップワードファイル

7

指定方法ラベル	ストップワード
0	前置詞
0	冠 詞
1	be
1	have
・	・
・	・
・	・

第 14 図

単語頻度テーブル

407

出 現 単 語	出 現 頻 度
translation	4
HICATS	3
document	2
system	2
・	・
・	・
・	・

第 16 図

文得点テーブル
408

文番号	重要語数	単語数	得　点
B 1	0	8	0
B 2	2	1 1	0.1 8 2
B 3	0	1 4	0
B 4	1	1 4	0.0 7 1
・ ・ ・	・ ・ ・	・ ・ ・	・ ・ ・

第 17 図

抄録文テーブル
409

抄録文番号	文　番　号
A 1	B 2
A 2	B 5
A 3	B 7
・ ・ ・	・ ・ ・

第 18 図

接続語句ファイル
8

接 続 語 句	拡　張　幅
but	− 1
however	− 1
as follows	− 1
therefore	− 1
・ ・ ・	・ ・ ・

第 19 図

対応範囲テーブル
410

抄録文番号	対応範囲開始位置	対応範囲終了位置
A 1	B 1	B 3
A 2	B 4	B 6
A 3	B 6	B 7
・ ・ ・	・ ・ ・	・ ・ ・

第 20 図

表示単語テーブル
411

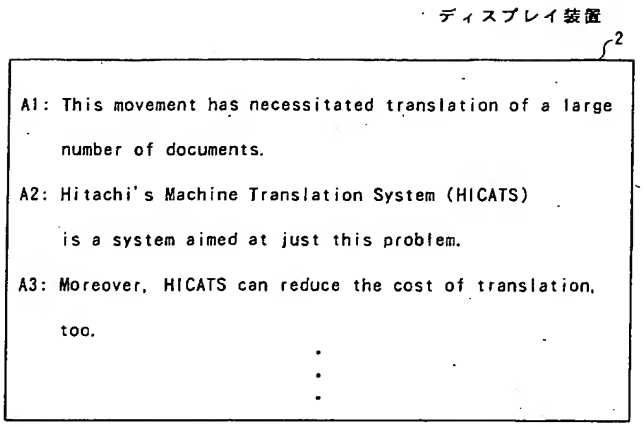
表 示 単 語	出 現 頻 度
HICATS	2
system	2
translation	2
・ ・ ・	・ ・ ・

第 21 図

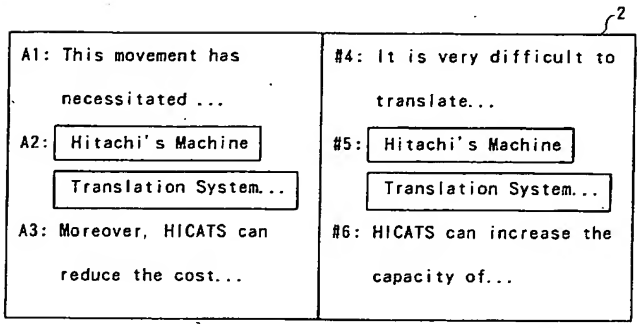
キーワードテーブル
412

キーワード
HICATS
translation
cost
・ ・ ・

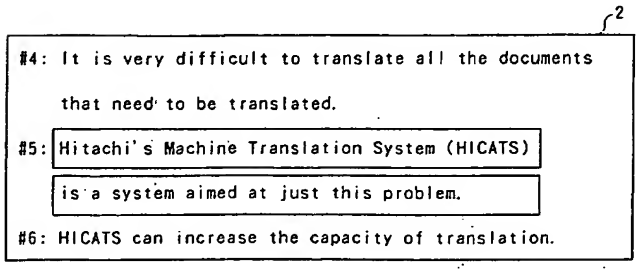
第 22 図



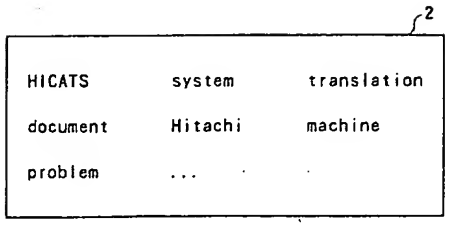
第 23 図



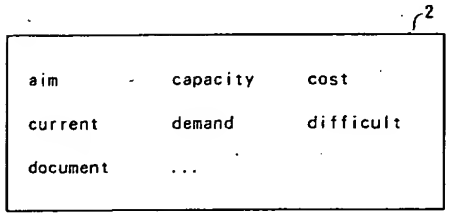
第 24 図



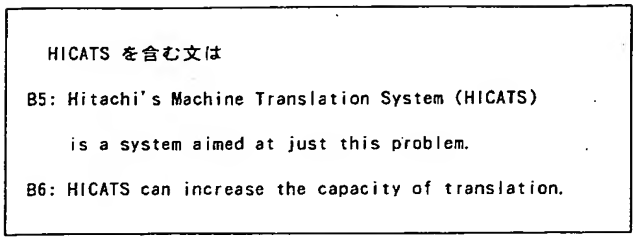
第 25(a) 図



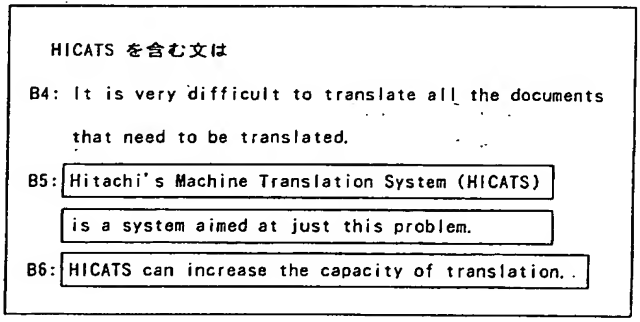
第 25(b) 図



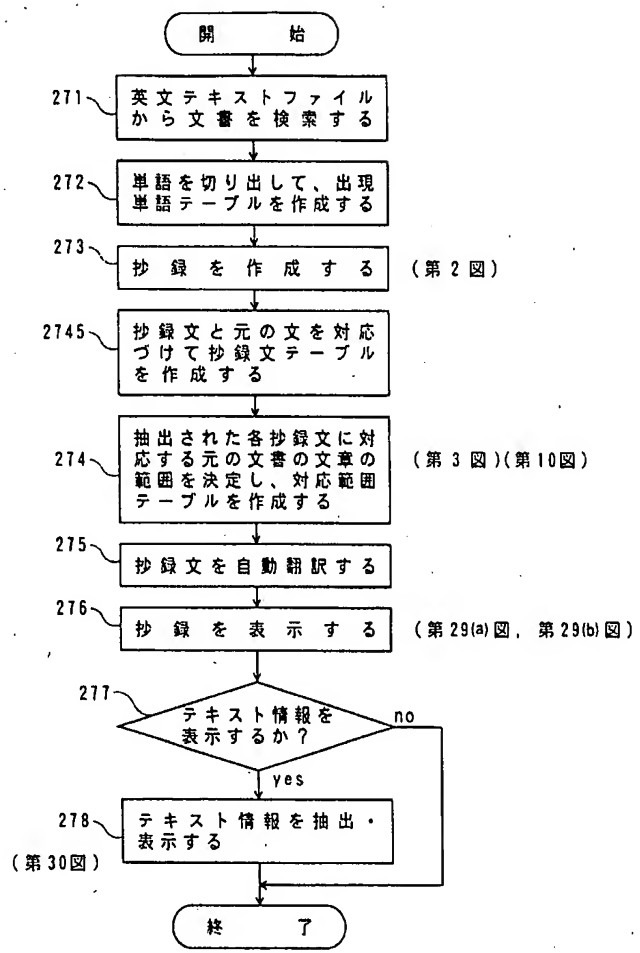
第 26(a) 図



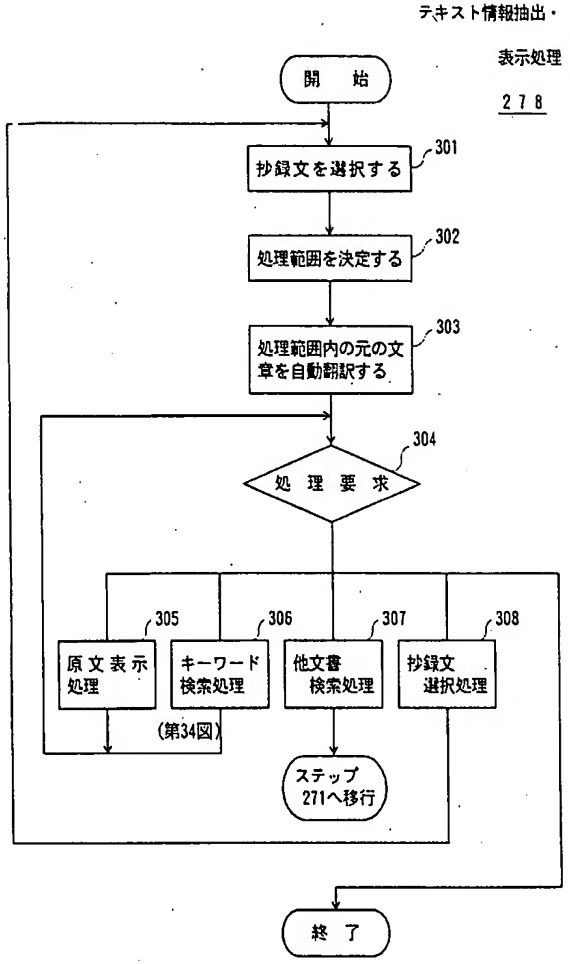
第 26(b) 図



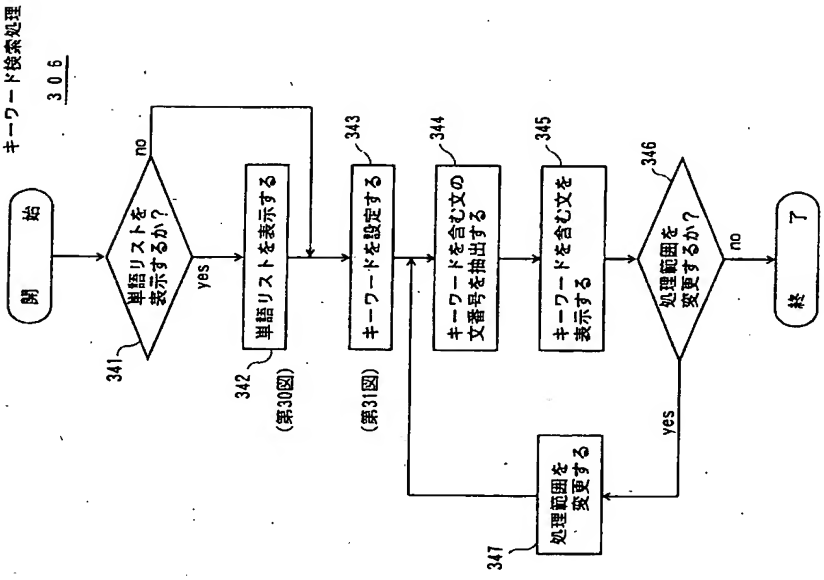
第 27 図


開　始
271
英文テキストファイルから文書を検索する
272
単語を切り出して、出現単語テーブルを作成する
273
抄録を作成する
（第2図）
2745
抄録文と元の文を対応づけて抄録文テーブルを作成する
274
抽出された各抄録文に対応する元の文書の文章の範囲を決定し、対応範囲テーブルを作成する
（第3図）（第10図）
275
抄録文を自動翻訳する
276
抄録を表示する
（第29(a)図，第29(b)図）
277
テキスト情報を表示するか？
no
yes
278
テキスト情報を抽出・表示する
（第30図）
終　了

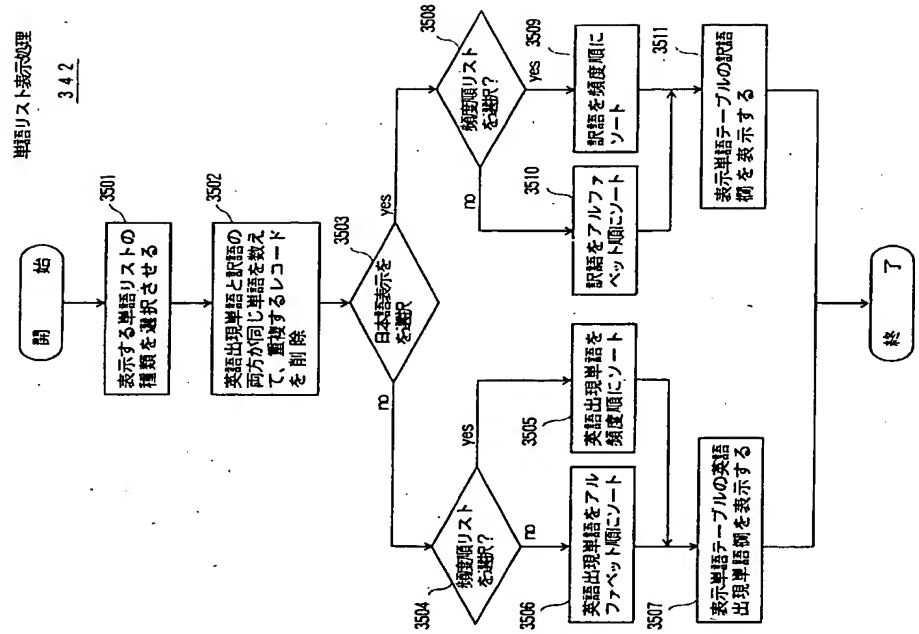
第 28 図


テキスト情報抽出・表示処理
278
開　始
抄録文を選択する
301
処理範囲を決定する
302
処理範囲内の元の文章を自動翻訳する
303
304
処理要求
305
原文表示処理
306
キーワード検索処理
307
他文書検索処理
308
抄録文選択処理
（第34図）
ステップ271へ移行
終　了

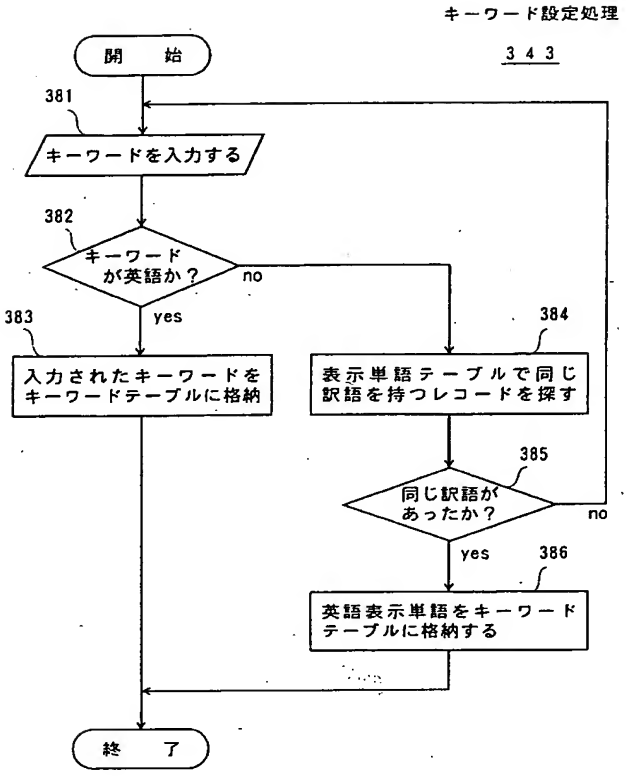
第 29 図


キーワード検索処理
306
開　始
単語リストを表示するか?
341
no
yes
単語リストを表示する
342
(第30図)
キーワードを設定する
343
(第31図)
キーワードを含む文の文番号を抽出する
344
キーワードを含む文を表示する
345
処理範囲を変更するか?
346
no
yes
処理範囲を変更する
347
終　了

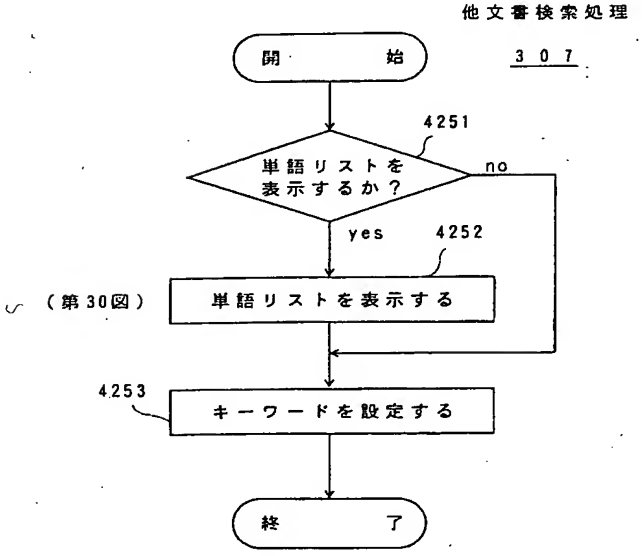
第 30 図


単語リスト表示処理
342
開　始
表示する単語リストの種類を選択させる
3501
英語出現単語と訳語の両方が同じ単語を除いて、重複するレコードを削除
3502
日本語表示を選択
3503
yes
no
頻度順リストを選択?
3504
yes
no
英語出現単語を頻度順にソート
3505
英語出現単語をアルファベット順にソート
3506
表示単語テーブルの英語出現単語を表示する
3507
頻度順リストを選択?
3508
yes
no
訳語を頻度順にソート
3509
訳語をアルファベット順にソート
3510
表示単語テーブルの訳語を表示する
3511
終　了

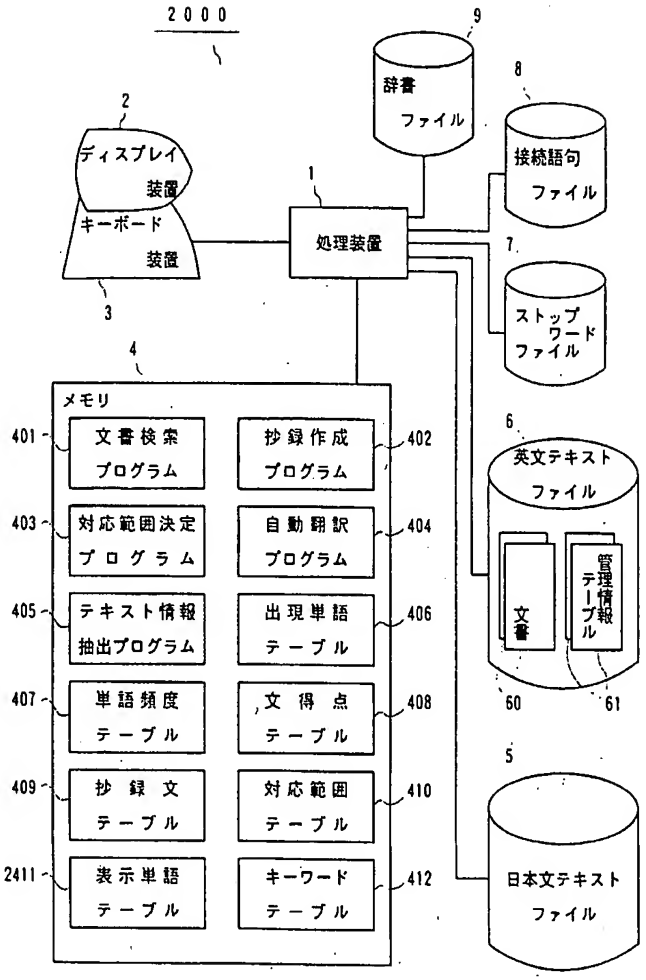
第 31 図


キーワード設定処理
343
開　始
381
キーワードを入力する
382
キーワードが英語か？
no
yes
383
入力されたキーワードをキーワードテーブルに格納
384
表示単語テーブルで同じ訳語を持つレコードを探す
385
同じ訳語があったか？
no
yes
386
英語表示単語をキーワードテーブルに格納する
終　了

第 32 図


他文書検索処理
307
開　始
4251
単語リストを表示するか？
no
yes
4252
（第30図）
単語リストを表示する
4253
キーワードを設定する
終　了

第 33 図



第34(a)図

文書
60

文番号	英文データ
B1	Japan's rate in …
B2	This movement …
B3	The number of …
	・ ・ ・

第34(b)図

日本テキストファイル
5

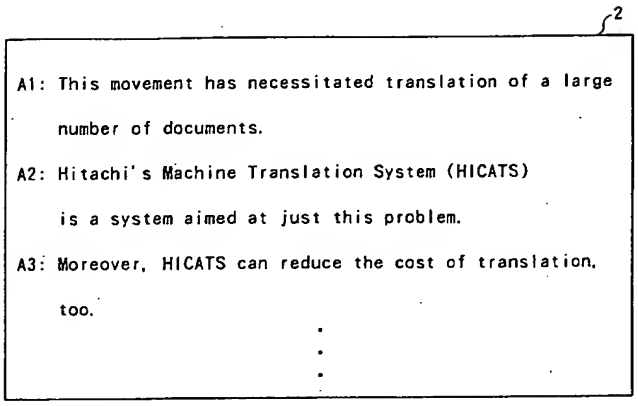
文番号	日本文データ
B1	
B2	この動きはたくさんの…
B3	
	・ ・ ・

第 35 図

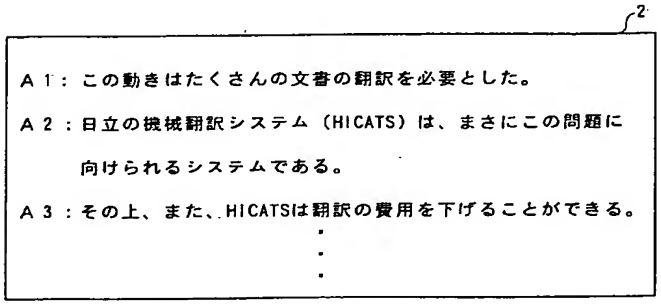
表示単語テーブル
2411

英語出現単語	訳　語	出現頻度
difficult	難しい	1
translate	翻訳する	2
document	ドキュメント	2
・ ・ ・	・ ・ ・	・ ・ ・

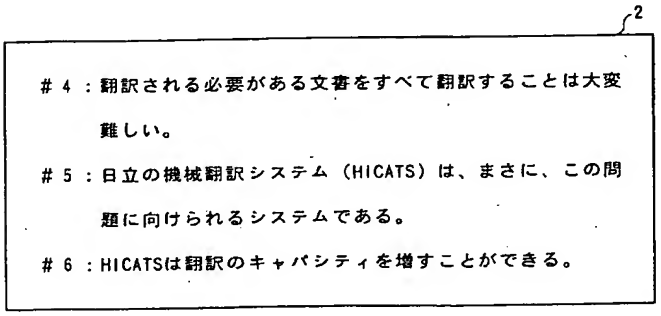
第 36(a) 図



第 36(b) 図



第 37(a) 図



第 37(b) 図

2

A 1 ：この動きはたくさんの文書の翻訳を必要とした。 A 2 ：日立の機械翻訳システム(HICATS)はまさにこの… A 3 ：その上、また、HICATSは翻訳の費用を下げること……	# 4 ：翻訳される必要がある文書をすべて翻訳する… # 5 ：日立の機械翻訳システム(HICATS)はまさにこの… # 6 ：HICATSは翻訳のキャパシティを増すことができる。

第 38(a) 図

2

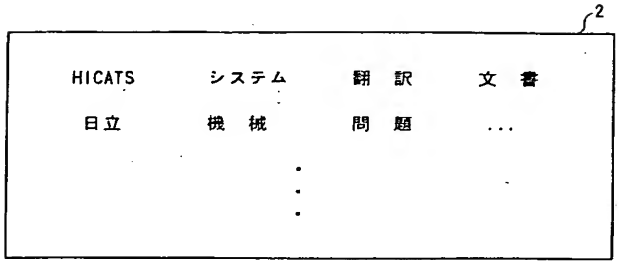
HICATS	system	translation	document
Hitachi	machine	problem	...

第 38(b) 図

2

aim	capacity	cost	current
demand	difficult	document	...

第 39(a) 図



第 39(b) 図